Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 65 (2015) 115 – 124
Procedia Computer Science

International Conference on Communication, Management and Information Technology (ICCMIT 2015)

Performance Evaluation of the PWSSIM Metric for HEVC and H.264

Jean Felipe F. de Oliveira[a], Carlos Danilo M. Regis[b], Marcelo Sampaio de Alencar[c]*

[a]*TPV Technology Limited, Rod. Anhanguera, Km 49 - Sitio Moenda - Tijuco Preto - Jundiai/SP - CEP 13205-700 – Brasil*
[b]*IFPB Av. 1° de Maio, Jaguaribe - João Pessoa/PB – Brasil*
[c]*Iecom – R. Aprígio Veloso 882, Bodocongó - Campina Grande/PB – Brasil*

Abstract

Considering the current requirements of available bandwidth due to the increase of video transmissions, a new video encoding standard was released by the ISO/IEC MPEG and ITU-T VCEG groups. The main feature of the High Efficiency Video Coding (HEVC) is to enable improved compression performance relative to existing standards in the range of 50% bit-rate reduction for equal perceptual video quality. Although the PSNR metric can show good results, it is not sensitive to several perceptual distortions which can perceive by a human being. The recently-proposed PW-SSIM metric incorporates the spatial information complexity to the SSIM metric. The HEVC encoding process maps some visual characteristics in order to avoid coding of redundant information. This paper comparatively evaluates the performance of the reference implementation of the HEVC and H.264 using the PW-SSIM metric, investigating its ability of mapping the aforementioned HEVC characteristic of incorporation of human vision heuristics.

Peer-review under responsibility of Universal Society for Applied Research
Keywords: Video coding; HEVC; H.264; Video quality evaluation; Objective metrics;

* Corresponding author. Tel.: +5511969230043; fax:+551145318414.
E-mail address: jean.oliveira@tpv-tech.com

Peer-review under responsibility of Universal Society for Applied Research
doi:10.1016/j.procs.2015.09.088

1. Introduction

The screen resolution and the exhibition resources of a wide range of consumer electronics products, from television to mobile devices, have significantly improved in the past recent years. This fact should further increase with the development of new technologies related to Ultra High Definition Television (UHD TV), with resolutions up to 7680 × 4320[1]. With the adoption of high-resolution devices, both wired and wireless (smartphones and tablets), the network operators will face new challenges in providing services with sufficient bandwidth to satisfy the consumer demand.

These new challenges can be met, at least partially, by the improvement of the compression rate, thereby reducing the bandwidth requirements of the current most used coding standard, the H.264 Advanced Video Coding. The new standard, the High Efficiency Video Coding (HEVC), aims to solve this problem by providing a 50% increase in compression efficiency over the H.264/AVC standard, maintaining an equivalent level of perceived visual quality.

The HEVC video coding standard is the latest joint effort between the standardization organizations ITU-T VCEG (Video Coding Experts Group) and ISO/IEC MPEG (Moving Picture Experts Group) in a partnership known as JCT-VC (Joint Collaborative Team on Video Coding.) The first edition of the HEVC standard was published in January 2013 in two rules with same contents disclosed by the organizations involved.

The HEVC standard was designed to meet all applications supported by the H.264/MPEG-4 AVC and especially with two specific requirements: encoding high and ultra-high resolution video and a better usage of parallel processing architectures[2,11].

It was shown that the incorporation of the perceptible spatial information (SI) to the structural similarity index (SSIM) metric, as an estimation of visual selective attention, led to an improvement of the quality evaluation of H.264 and MPEG-2 encoded videos with gains of around 21% compared to the standard SSIM metric[3,4]. This improvement is explained by the fact that the SI is less sensitive than the PSNR and SSIM when comparing the distortions introduced by some process of H.264 and MPEG-2 encoding, e.g. blur.

The purpose of this article is to comparatively evaluate the coding efficiency of the reference implementation of the HEVC and the H.264 encoder using the recently proposed PW-SSIM metric[5], analyzing if the new enhancements of the HEVC, regarding the exploitation of spatial information, can be reflected on the PW-SSIM results. The PW-SSIM is also compared with the behavior of the embedded PSNR of each encoder.

2. HEVC Architecture and Main Features

The HEVC standard is designed to achieve several goals, including the high coding efficiency, ease of integration of transport systems and resilience in loss of information, as well as adaptability to hardware architectures with parallel processing. In the following sections the main points of the architecture of HEVC standard are described, highlighting the new features, and the specific operations for the production of a valid HEVC bitstream.

2.1. The Coding Tree Units Structures

The main motivation to use the block-based partitioning in video or image compression is the its ability to code each block with a specific configuration chosen within a set of pre-defined parameters, taking into account that, in general, a single model can not efficiently map the properties of a complete picture. The Quadtree data structure allows partitioning of an image into blocks of variable size and is therefore suitable to optimize the tradeoff between the accuracy of the model (given by a distortion D) and the cost of the coding model, typically measured in bitrate R. Tree algorithms that optimize the lagrangian functional were developed motivating the design of compression algorithms[2,6]. The HEVC encoder/decoder uses the Quadtree partitioning approach using syntactic elements which can store the information of different types of block subdivisions. These syntactic elements are also used as parameters of the steps of transformation and prediction on the encoder.

Also, if a data block is divided into four blocks, all the child blocks are typically coded separately, even if two or three of them share the same encoding parameters. However, the suboptimal properties of the initial Quadtree subdivision can be substantially improved by merging of nodes (blocks) belonging to potentially different parents.

It should be noted that an algorithm of merging of spatially neighboring blocks is conceptually similar to the spatial prediction modes, for example, the spatial direct mode of the H.264/AVC standard. This mode also tries to reduce the coding cost using redundant motion parameters of neighboring blocks. However, the improvements compared to H.264/AVC, suggest that the concept of merging is superior to exploit these redundancies[2,7]. That argument is tested in this paper, evaluating whether this new spatial organization, which tends to be more aligned with the visual perception of quality, can produce better results with objective assessment techniques that maps subjective aspects of vision[11].

2.2. The Block Merging Algorithm

The purpose of block merging is to compensate the drawbacks of the initial sub-divisions of the Quadtree algorithm, reducing redundant sets of coding parameters to be transmitted[8]. Figure 1 illustrates a possible partitioning of a prediction block. This is achieved by creating regions composed by neighboring prediction blocks, sharing identical motion informations, which will be signaled once per region. This way, each region should have at least one prediction block not fused to a neighbor and provide motion information to others.

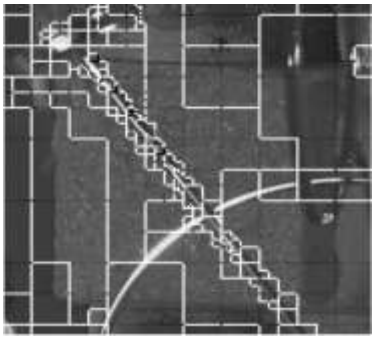

Figure 1 Example of the Block merging algorithm divisions.

2.3. The New Block Structure

The core of the coding layer of previous standards was the macroblock, composed by a block of 16×16 luminance samples and, for the usual case of a 4:2:0 chroma subsampling, two 8×8 blocks of chrominance samples. While the HEVC standard has, as a macroblock equivalent structure, a format known as Coding Tree Unit, which will be called, throughout this text, CTU. The CTU has a real-time selected size in the encoding process and may be larger or smaller than a traditional macroblock size. A CTU consists of a coding tree block (CTB) of luminance information, two chrominance CTBs and the associated syntactic elements. The size of a luminance CTB can be equal to 16×16 , 32×32 or 64×64 samples, with larger sizes typically providing better compression. Thus, the HEVC supports a partitioning of CTBs into smaller blocks using a tree structure based on the Quadtree algorithm.

2.4. Prediction Units and Prediction Blocks

The decision to encode the region of an image using inter or intra prediction is realized on the coding unit (CB). A partition structure of the prediction unit has its own root. Depending on the prediction decision, the luminance and chrominance CBs can be further divided into blocks for luminance and chrominance prediction. The HEVC encoding process can support a prediction block size ranging from 64×64 to 4×4.

2.5. Transform Units and Transform Blocks

The generated residuals from the prediction stage are encoded using entities known as Transform Blocks (TB). The luminance TB size can match the luminance CB size, but can be splitted into small TB. A domain-specific DCT is defined to calculate the transforms. An integer transform, derived from the DST, is specified to be applied at 4×4 blocks.

2.6. Advanced Motion Vector Prediction

A new technique called Advanced Motion Vector Partition (AMVP) was created on the HEVC standard. This technique consist in the inclusion of derivations of the most likely candidate vectors based on the motion information of adjacent prediction blocks and the reference frame for a fusion of motion vectors that inherits the vectors of spatial or temporal neighbors prediction blocks. And yet, comparing with the H264/AVC, new enhancements have been made to the inference techniques of skip vector and direct motion inference.

2.7. Motion Compensation

Similarly to the H.264/AVC, multiple reference frames are used. For each PB, one or two motion vectors may be transmitted as the results of one-way and bi-directional prediction modes, respectively. An operation of scaling and shifting, known as weighted prediction can be applied to the resulting video of the prediction and motion compensation process.

2.8. Intra Prediction

The HEVC intra prediction process supports 33 directional modes, besides the planar and DC modes. The H.264/AVC had specified eight directional modes. The selected intra prediction modes are encoded deriving the most probable mode based on the information of previously encoded neighboring prediction blocks.

2.9. Sample Adaptive Offset Filter

Compared to H.264/AVC, in which only deblocking filtering is applied in the process of reference image recover within the encoder, the HEVC standard specifies a new filter called Sample Adaptive Offset (SAO). The SAO introduces a nonlinear amplitude mapping in the inter prediction process after the deblocking filter. This technique aims to a better reconstruction of the amplitude of the original signal using a lookup table which is described by some additional parameters that can be determined by examining the histogram on the encoder side.

3. Objective Metrics for Video Quality Assessment

3.1. The Peak-to-Noise Signal Ratio (PSNR)

For the evaluation of video quality the PS NRYUV metric is one of the most used[7] and is defined as,

$$PSNR_{YUV} = \frac{6 \cdot PSNR_Y + PSNR_U + PSNR_V}{8}$$

in which the values of PS NRY , PS NRU e PS NRV are calculated using the Formula

$$PSNR = \frac{10 \cdot \log_{10}(2^B - 1)^2}{MSE}$$

in which B = 8 is the number of bits per sample of the video signal to be encoded and MSE (Mean Squared Error) is the value of the sum of the squared differences divided by the number of signal samples.

3.2. The Perceptual Weighting Structural Similarity Index (PWSSIM)

The PW-SSIM method uses the perceptual spatial information as a way of weighting the most visually important regions. This weighting is obtained as follows: first the magnitude of the gradient vectors of the original video is calculated using Sobel masks, then it generates a table in which the values of pixels are the magnitudes of the gradients. Then, this frame is partitioned into blocks of 8×8 pixels and the Spatial Information (SI) is calculated using the Equation

$$SI = \left(\frac{1}{N-1} \cdot \sum_{i=1}^{N} \left(\mu_s - S \right)^2 \right)^{\frac{1}{2}}$$

in which µS is the mean value of the gradient magnitude of a block and N is the number of pixels in the block. Based on this consideration, the value of S I was incorporated into the SSIM, leading to the model so called Structural Similarity Index with Perceptual Weighting, given by

$$PWSSIM(f,h) = \frac{\sum_{d=1}^{D} SSIM_d(f,h) \cdot SI_d}{\sum_{d=1}^{D} SI_d}.$$

4. Performance Evaluation

The results of an objective evaluation that compares the HEVC encoding performance with the H.262/MPEG-2 Video, H.263, MPEG-4 Visual, H.264/MPEG-4 Advanced Video Coding (AVC) capabilities are exhibited in [7]. These evaluations indicate that the bitrate was almost halved compared to the rates obtained with the H.264/AVC reference encoder (JM), considering the same subjective quality. Configured with the parameters for the LP (Low Profile), the HM encoder can achieve a reduction of 50% in the output bitrate output relative to the H.264[2].

The results of subjective evaluations shows a reduction of the average bitrate of an HEVC encoder over all sequences enconded with a setup that provides an equivalent subjective quality was estimated at 67% for the five full HD sequences 49% for the four WVGA, and 58% in average[9]. These initial measurements indicate that the HEVC satisfy its initial goal in terms of subjective quality gains. Although the number of test cases has been limited, because the same trend can be observed for all the different sequences, it is an indicative of the significant improvement in compression efficiency which is provided by HEVC.

Althought trustable, these subjective video quality evaluations, such as MOS (Mean Opinion Score), which tends to be the most reliable metric, have a quite complex implementation. This way, measures of automatic and repeatable objective quality as the Peak Signal to Noise Ratio (PSNR), the Structural Similarity Index (SSIM) and Perceived Quality Index are typically used when subjective measures are impractical or difficult to implement.

It was showed that the an HEVC encoder have an average bitrate saving of approximately 40% compared to H.264 and 70% to 80% when compared to MPEG-2 Main Profile[7]. This study used a disciplined comparison approach regarding the choice of tools and coding parameters (profiles and levels) to ensure a fair comparison between all the chosen encoders. Besides the average gain in compression efficiency, tests indicated some other important characteristics in comparison with other standards[10]. They are:

- The benefits in terms of subjective video quality seemed higher than what is suggested by the results of PSNR. This is the aspect investigated in this article.
- The gain in compression efficiency indicated to be greater for high-definition video than for videos with small resolutions.
- For the tested configurations, the gains were higher in videos configured for real-time communication than for entertainment (low delay) applications.
- The lower the bitrate, the larger the proportional gain in perceived quality.

5. V. RESULTS

5.1. Reference Videos

The academic research community, which deals with image and video processing, generally uses certain set of videos to allow the experiments to be repeated in other laboratories and compared by researchers of other laboratories results comparisons. The utilized videos in this work were: BasketballDrive(1920x1080) and Kimono1(1920x1080). In general, most videos had a variable number of frames. Thus, they were all reduced to 50 frames, since only spatial characteristics were observed in this study. This value was chosen to reduce the high computational load due to limitations in the computer´s configurations used.

5.2. Simulations

The main objective of this study is to comparatively evaluate the performance of the PWSSIM metric against the PSNR metric, confronting the results of the H.264 and HEVC encoding processes with the premise that the HEVC addresses more heuristics of human vision and should have better results than H.264. All videos were encoded and decoded by the reference codecs proposed by the committees responsible for the H.264 and HEVC, respectively. And for each decoded video, the PSNR and the PWSSIM was calculated.

This comparison was observed by varying three parameters: the output bitrate, the quantization parameter (QP) and video resolution. The quantization parameter is the element that controls all the quantization process used in the encoders. It is used to calculate the scaling matrix and quantization step, which is a value that determines the variation range during the quantization process. In the simulations, this value is pre-set to zero. The resolution defines the dimensions of the video and determines the amount of information to be encoded. Figure 3 shows a comparison of the file size of the encoded sequence of the BasketballDrive video file and, as promised, the H.264 file size exceeds 50% of the HEVC file size, especially for higher resolutions.

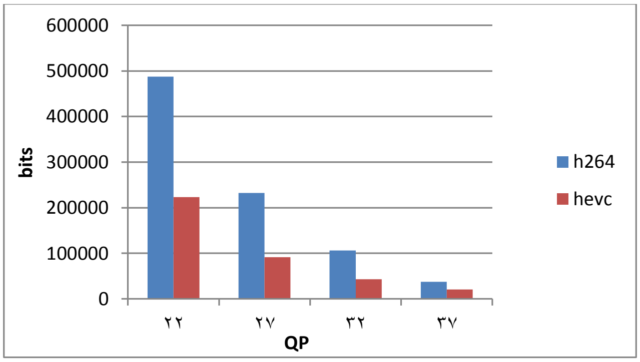

Figure 2: Comparison between the sizes of the resulting encoded file.

Figure 3 shows the results of the PWSSIM metric, while the quantization parameter is increased. The choice of these parameters was made according to the reviewed literature. An increase on the quantization parameter implies a decrease on the encoded sequence size. Figure 4 shows the tendency of the PW-SSIM and the PSNR metrics of a HEVC encoded videos with different quantization parameters. Figure 3 shows that the PWSSIM results for H.264 are slightly better for all values of the tested parameter encoding. However, the size of the encoded file, expressed in Figure 2, shows large gains using the HEVC encoder with almost the same objective quality measurement.

The simulations concerning the variation of the resolution were performed using two video sequences of different characteristics in relation to spatial complexity and the amount of movement during the video. The Kimono1 sequence and the sequence BasketballDrive were sub-sampled for the following resolutions: 1280×720, 960×576, 352×288 (CIF) e 176×144 (QCIF).

Each resulting YUV video was encoded and decoded by the same codec and the PSNR and PWSSIM were measured for each case. The results were plotted in the graphs shown in Figures 5 and 6 for Kimono1 sequence. The simulations concerning the variation of the resolution were performed using the Kimono1 sequence, which was originally captured in 1920 × 1080 and sub-sampled for the following resolutions: 1280×720, 960×576, 352×288 (CIF) e 176×144 (QCIF).

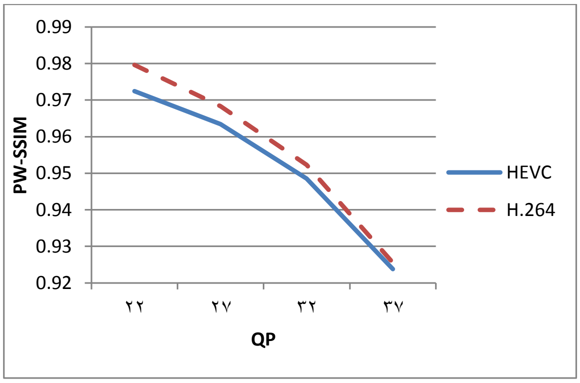

Figure 3. Comparison of the quality of the resulting encoded file.

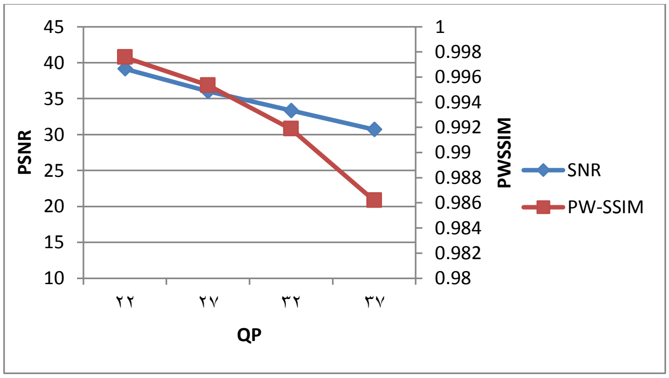

Figure 4. Comparison of the quality of the resulting encoded file.

Each resulting YUV video was encoded and decoded by the same codec and the PSNR and PWSSIM were measured for each case.The results were plotted in the graphs shown in Figures 5 and 6 for Kimono1. Figures 7 and 8 shows the behavior of the PWSSIM and PSNR when increasing the target output bitrate on the encoding process of the Kimono1 FHD (Full HD) sequence. An analysis of the results of Figure 7, indicates that the PSNR metric does not addresses the distortions on the spatial information for HEVC encoded videos, as the PWSSIM metric does. Figure 8 indicates that for H.264 videos, the PSNR and PWSSIM seems to have a similar behaviour.

6. Conclusions

The objective of this research was to analyze the performance of an objective metric designed for digital video evaluation, the recently proposed PWSSIM. This new metric combines heuristics of human vision with the index of structural similarity (SSIM) and, arguably, managed to obtain a good correlation with subjective metrics.

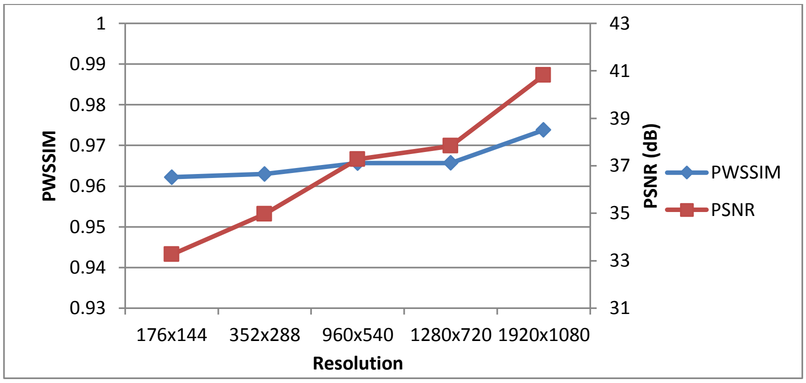

Figure 5: Comparison of the PSNR and the PWSSIM tendencies for a HEVC encoded video.

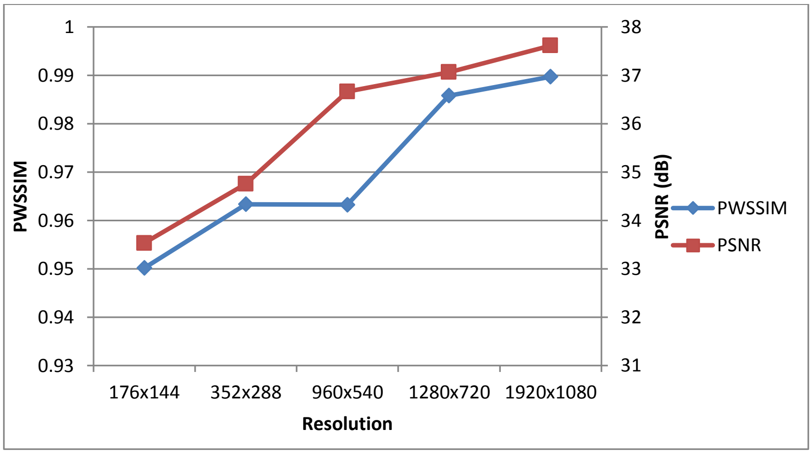

Figure 6: Comparison of the PSNR and the PWSSIM tendencies for a H.264 encoded video.

The results of the application of this metric were observed for H.264 and HEVC encoded videos, which were compared to the results of the PSNR metric embedded into each encoder. Considering the characteristic of the HEVC encoder to recognize and incorporate a number of human vision characteristics, such as the tolerance to spatial redundancy, especially in the process of image segmentation to form blocks of variable sizes, a different tendency for the HEVC was expected in the comparison with the H.264 encoding process. Indeed, the behaviour of the curves obtained using the PWSSIM showed a non-linear tendency for the H.264, which can be especially noticed when comparing the metrics performance against the target output bitrate from Figures 7 and 8.

For the continuation of this study, a comparison with subjective ratings and the calculation of correlation will be realized.

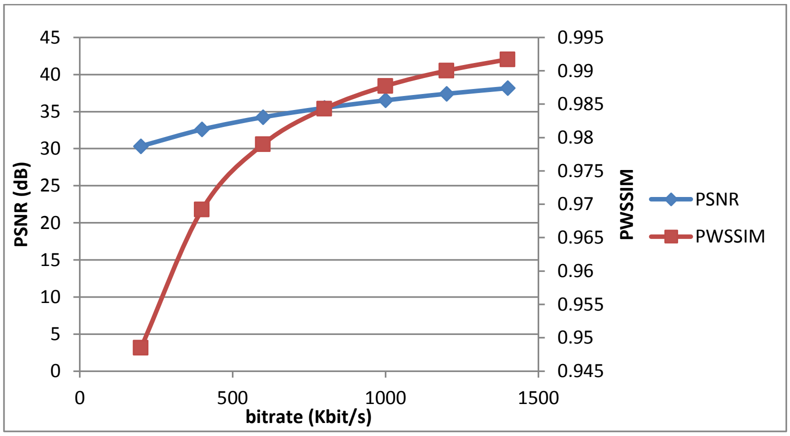

Figure. 7: PSNR and PWSSIM plotted against HEVC encoded video bitrate

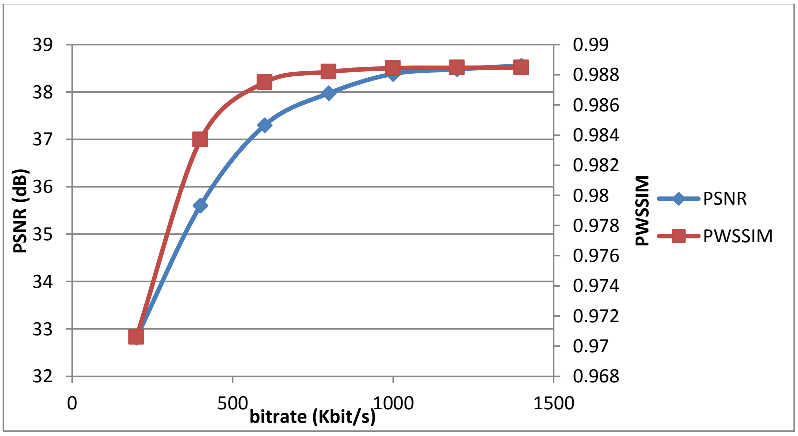

Fig. 8: PSNR and PWSSIM plotted against H.264 encoded video bitrate

7. Future Works

A research opportunity is the real-time optimization of the parameterization of the main modules of the encoding process based on the extraction of information such as the spatial complexity and temporal complexity using bio-inspired algorithms. Objective evaluation of HEVC 3D videos is also a candidate subject for future studies, applying the same methodology of this study.

8. Acknowledgement

The authors would like to thank the Copele (UFCG), the IFPB, the Institute for Advanced Studies in Communications (Iecom) and the TPV-Tech company for the support to this research.

References

[1] J. Nightingale, Q. Wang, and C. Grecos, “HEVStream: A Framework for Streaming and Evaluation of High Efficiency Video Coding (HEVC) Content in Loss-prone Networks,” IEEE Trans. Consumer Electronics, vol. 58, no. 2, pp. 404–412, 2012.

[2] J. Vanne, M. Viitanen, T. D. Hamalainen, and H. A., “Comparative Rate- Distortion-Complexity Analysis of HEVC and AVC Video Codecs,” Circuits and Systems for Video Technology, IEEE Transactions on, vol. 22, no. 12, pp. 1885–1898, 2012.

[3] M. S. A. Carlos D. M. Regis, “Métrica de Avaliação Objetiva de Vídeo Usando a Informação Espacial , a Temporal e a Disparidade,” Ph.D. dissertation, Federal University of Campina Grande – UFCG, 2013.

[4] C. D. M. Regis, I. P. Oliveira, and M. S. Alencar, “Combining Spatial and Temporal Analysis into an Objective Video Quality Metric,” vol. X,no. X, pp. 1–5.

[5] I. Carlos D. M. Regis, Student Member, IEEE, José V. M. Cardoso, and Marcelo S. Alencar, Senior Member, “Performance of the Structural Similarity Index Using the Spatial Perceptual Information for Video Quality Assessment,” in Journal of Eletronic Imaging – SPIE, no. 4, pp. 1–3.

[6] X. L. Ohm, M. Wien, and Jens-Rainer, “Rate-Complexity-Distortion Evaluation for Hybrid Video Coding” in Multimedia and Expo (ICME), IEEE International Conference on, Suntec City, 2010, pp. 685–690.

[7] J.-r. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, S. Member, and T. Wiegand, “Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC),” IEEE Trans. Circuits Syst. Video Techn., vol. 22, no. 12, pp. 1669–1684, 2012.

[8] P. H. Wiegand, S. Oudin, B. B. and Detlev Marpe, M. O. Bici, K. Ugur, J. Jung, G. Clare, and Thomas, “Block Merging for Quadtree-Based Partitioning in HEVC,” IEEE Trans. Circuits Syst. Video Techn., vol. 22, no. 12, pp. 1720–1731, 2012.

[9] J.-R. O. V. Baroncini, “Report on preliminary subjective testing of HEVC compression capability,” ITU-T/ISO/IEC Joint Collaborative Team on Video Coding document JCTVC-H1004,, 2012.

[10] J.-R. O. Andrew Segall, Jill Boyce, Gary J. Sullivan, “Report of Results of the Joint Call for Proposals on Scalable High Efficiency Video Coding (SHVC),” Shanghai, China.

[11] Oliveira, J. F. F., Alencar, M. S., Padrão HEVC – Novas Tecnologias para Aplicações de Elevadas Taxas de Compressão de Vídeo. Revista de Tecnologia da Informação e Comunicação Vol 4. Num 2, Outubro 2014.

[12] L. Song, X. Tang, W. Zhang, X. Yang, P. Xia, The SJTU 4K Video Sequence Dataset, the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), Klagenfurt, Austria, July 3rd-5th, 2013